Formative and summative assessment – a harmonious relationship?

Wynne Harlen

Paper given at the ASF Seminar, January 2005

This paper[1] explores the extent to which assessment information can be used for both summative and formative purposes, without the use for one purpose endangering the effectiveness of use for the other. Some of those involved in developing assessment have argued that the distinction is not helpful and that we should simply strive for 'good assessment'. Good formative assessment will support good judgements by teachers about student progress and levels of attainment (AifL, 2004) and good summative assessment will provide feedback that can be used to help learning. Maxwell (2004) describes progressive assessment, which we consider below, as blurring the boundary between formative and summative assessment. However, it remains the case that formative and summative are different purposes of assessment and while the same information may be used for both, it is necessary to ensure that the information is used in ways that serve these purposes. It seems that, under current arrangements, in practice information is gathered initially with one of these purposes in mind and may or may not be used for the other. These are arguments to return to after looking at these current practices and considering the possibility of collecting information designed for both purposes.

Using summative assessment to help learning

Feedback is a key factor in formative assessment, for only by indicating to the students what is needed for their next steps in learning can the teacher bring about the intended engagement in learning. He or she cannot do the learning for their students. However, the teacher makes use of feedback to adjust teaching, making materials and opportunities for the learning available and, most importantly, making clear the purposes and goals of the work.

Some examples of using assessment in this way are provided by Maxwell (2004) and Black et al (2003). Maxwell describes the approach to assessment used in the Senior Certificate in Queensland, in which evidence is collected over time in a student portfolio, as 'progressive assessment'. He states that

> All progressive assessment necessarily involves feedback to the student about the quality of their performance. This can be expressed in terms of the student's progress towards desired learning outcomes and suggested steps for further development and improvement….
> For this approach to work, it is necessary to express the learning expectations in terms of common dimensions of learning (criteria). Then there can be discussion about whether the student is on-target with respect to the learning expectations and what needs to be done to improve performance on future assessment where the same dimensions appear.

[1] Part of a paper entitled 'Teachers' summative practices and assessment for learning – tensions and synergies' to be published in a special edition of *The Curriculum Journal* edited by Paul Black, 2005.

> As the student builds up the portfolio of evidence of their performance, earlier assessment may be superseded by later assessment covering the same underlying dimensions of learning. The aim is to report 'where the student got to' in their learning journey, not where they started or where they were on the average across the whole course.
> (Maxwell 2004, pages 2 and 3)

The identification of goals and assessment criteria in terms of a 'common dimension of learning' is, as Maxwell states, central to this approach. Further, descriptions of these dimensions of learning need to be detailed to be capable of giving guidance, yet not so prescriptive as so infringe teachers' ownership of the curriculum. As the research shows (ASF, 2004), the dependability of assessment is enhanced when teachers have a thorough understanding of the goals and of the nature of progression towards them. In Queensland, this is facilitated on the one hand by schools being able to make decisions about their own work plan and, on the other hand, by teachers' regular participation in the process of moderation. Time and respect for the professionalism of teachers (Cumming and Maxwell, 2004) are also important. These are clearly essential factors when teachers' assessment has outcomes with high stakes for individual students. However, a significant feature of the Queensland system is that the assessment of students in the Senior Certificate is detached from school and teacher accountability procedures.

Black *et al* (2003) include the formative use of summative assessment as one of four practices that teachers found effective ways of implementing formative assessment (the others being questioning, feedback by marking and student peer-and self-assessment). These practices were all devised or elaborated by teachers as they strove, working with the researchers, to make changes in their classrooms so that assessment was used to help learning. In relation to the formative use of summative tests, the teachers devised three main ways of using classroom tests, beyond just assessing attainment, to develop students' understanding. The first of there involved helping students to prepare for tests by reviewing their work and screening past test questions to identify areas of insecure understanding. This reflection on their areas of weakness enabled them to focus their revision. The second innovation was to ask students to set test questions and devise marking schemes. This helped them 'both to understand the assessment process and to focus further efforts for improvement'. (Black *et al*, 2003, p 54). The third change was for the teachers to use the outcome of tests diagnostically and to involve students in marked each other's tests, in some cases after devising the mark scheme. This has some similarity to the approach reported by Carter (1997), which she called 'test analysis'. In this the teacher returned test papers to students after indicating where there were errors, but leaving the students to find and correct these errors. The students' final mark reflected their response to the test analysis as well as the initial answers. Carter described this as shifting the responsibility for learning to the students, who were encouraged to work together to identify and correct their errors.

These approaches are ones that teachers can use in the context of classroom tests over which they have complete control. Black *et al* (1993) noted that when external tests are involved, the process can move 'from developing understanding to "teaching to the test". More generally, the pressures exerted by current external testing and assessment requirements are not fully consistent with good formative practices'. (Black *et al*, 2003, p56). These teachers used their creativity to graft formative value

on to summative procedures. A more fundamental change is needed if assessment is to be designed to serve both purposes from the start.

There is the potential for such change in the use of computers for assessment, which provide the opportunity for assessment to serve both formative and summative purposes. In the majority of studies of the use of ICT for assessment of creative and critical thinking, reviewed by Harlen and Deakin Crick (2003), the assessment was intended to help development of understanding and skills as well as to assess the attainment in understanding and skills. The effectiveness of computer programs for both these purposes was demonstrated by those studies where computer-based assessment was compared with assessment by paper and pencil (Kumar *et al*, 1993; Jackson, 1989). The mechanism for the formative impact was the feedback that students received from the program. In some cases this was no more than reflecting back to the students the moves or links they made between concepts or variables as they attempted to solve a problem. In others (eg Osmundson *et al*, 1999) the feedback was in providing a 'score' for a concept map that they created on the screen by dragging concepts and links. The score compared the students' maps with an 'expert map' and required a much greater degree of analysis than could be provided in any other way. In other studies (Schacter *et al*, 1997) the computer program used a record of all mouse clicks in order to provide feedback to the students and teacher information about the processes used in reaching a solution. Schacter *et al* (1997) referred to this as 'bridging the gap between testing and instruction'.

In order for assessment to have a formative purpose it is necessary to be able to report not only the students' final performance, but also what processes students need to improve in order to raise their performance. The collection of information about processes, even if feasible in a non-computer-based assessment is immensely time-consuming and would not be a realistic approach to meeting the need for information for improving learning. The use of computers makes this information available, in some cases instantly, so that it provides feedback for the learner and the teacher that can be used both formatively as well as summatively. In these cases the process of assessment itself begins to impact on performance; teaching and assessment begin to coalesce. Factors identified as values of using computers for learning then become equally factors of value for assessment. These include: speed of processing, which supports speed of learning; elements of motivation such as confidence, autonomy, self-regulation and enthusiasm, which support concentration and effort; ease of making revisions and improved presentation which support quality of writing and other products; and information handling and organisation, which supports understanding (NCET, 1994).

Using formative assessment information for summative assessment

The approaches discussed above are linked to summative assessment as an occasional, if frequent, event. In between classroom tests, whether administered by computer or otherwise, there are innumerable other classroom events in which teachers gather information about the students by observing, questioning, listening to informal discussions among students, by reviewing written work and by using students' self-assessment (Harlen and James, 1997). In formative assessment this information may be used immediately to provide students with help or it may be stored and used to plan learning opportunities at a later stage. The information gathered in this way is

often inconclusive and may be contradictory, for what students can do is likely to be influenced by the particular context. This variation, which would be a problem for summative assessment, is useful information for formative purposes, suggesting the contexts in which students can be helped to develop their ideas and skills. By definition, information gathered at this level of detail relates to all the aspects of students' learning. It is valuable information that is well suited to deciding next steps for individual learners or groups. An important question is: can this rich but sometimes inconsistent information be used for summative assessment purposes as well as for formative assessment, for which it is so well suited? If not, then separate summative assessment will be necessary.

A positive answer to this question was given by Harlen and James (1997) who proposed that both purposes can be served providing that a distinction is made between the *evidence* and the *interpretation* of the evidence. For formative assessment the evidence is interpreted in relation to the progress of a student towards the goals of a particular piece of work, next steps being decided according to where a student has reached. The interpretation is in terms of what to do to help further learning not what level or grade a student has reached. For this purpose it is important for teachers to have a view of progression in relation to the understanding and skills they are aiming for their students to achieve. The course of progression can be usefully expressed in terms of indicators, which both serve the purpose of focusing attention on relevant aspects of students' behaviour and enable teachers to see where students are in development. An example of indicators for the development of observation and using information sources in the context of science at the primary level is given in Box 3.

Box 3 Example of developmental indicators

Things students do that are indicators of gathering information by observing and using information sources:

1. identify obvious differences and similarities between objects and materials
2. make use of several senses in exploring objects or materials
3. identify relevant differences of detail between objects or materials and identify points of similarity between objects where differences are more obvious than similarities
4. use their senses appropriately and extend the range of sight using a hand lens or microscope as necessary
5. take an adequate series of observations to answer the question or test the prediction being investigated
6. take steps to ensure that the results obtained are as accurate as they can reasonably be and repeat observations
7. regularly and spontaneously use printed and electronic information sources to check or supplement their investigations.

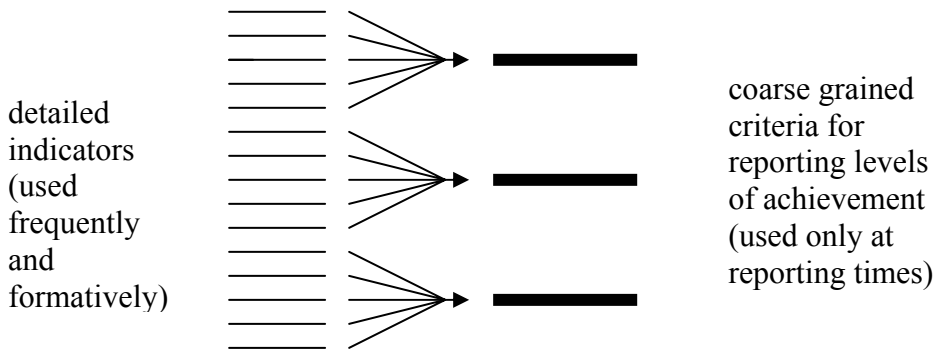
(from Harlen and Qualter, 2004)

These indicators have been developed from what is known about progression from research and practice, but they are not by any means definitive. It is not likely that there is an exact and invariable sequence that applies to every student, but is it helpful to have a rough idea. Examples of similar lists have been published in Australia (Masters and Forster, 1996) and developed in California (the Berkeley Evaluation and Assessment Research (BEAR) assessment system (Wilson, 1990; Wilson *et al* 2004).

In these lists, the earlier statements indicate understanding, skills or attitudes that are *likely* to be developed before the ones later in the list. There are no 'levels', grades, or stages suggested: just a sequence expected for students in a particular age range (in the example, in primary and early secondary school years). For formative assessment it is not necessary to tie indicators to grade level expectation – all that is required is to see where students are and what is the next step in their further progress.

For summative purposes, of course, common criteria need to be applied and achievement is generally summarised in terms of levels or grades that must have the same meaning for all students. This means that if the information already gathered and used formatively is to be used for summative assessment it must be reviewed against the broader criteria that define levels or grades. This process is one of finding the 'best fit' between the evidence gathered about each students and one of the reporting levels. In this process the change over time can be taken into account so that, as in the Queensland portfolio assessment, preference is given to evidence that shows progress during the period covered by the summative assessment. This process is similar to the one teachers are advised to use in arriving at their teachers' assessment for reporting at the end of key stages in the National Curriculum Assessment. The difference is that in the approach suggested here teachers have gathered information in ways suggested above (incorporating the key features of formative assessment) over the whole period of students' learning and used it to help students with their learning.

Figure 1 Formative and summative assessment using the same evidence but different criteria



The detailed indicators will map onto the broader criteria, as suggested in Figure 1. The mapping will smooth out any misplacement of the detailed indicators. But it is important *not* to see this mapping as a summation of judgements about each indicator. Instead the evidence is *re-evaluated* against the broader reporting criteria.

Conclusion

What has the research evidence reviewed in the ASF Working Paper 2 (ASF 2004 a) and the arguments presented here to say in relation to the questions of whether teachers' summative assessment and assessment for learning need to be considered as distinct from each other or how they can be harmonised? There seems to be value in maintaining the distinction between formative and summative *purposes* of assessment

whilst seeking synergy in relation to the *processes* of assessment. These different purposes are real. One can conduct the same assessment and use it for different purposes just as one can travel between two places for different purposes. As the purpose is the basis for evaluating the success of the journey, so the purpose of assessment enables us to evaluate whether or not the purpose has been achieved. If we fuse, or confuse, formative and summative purposes, experience strongly suggests that 'good assessment' will mean good assessment of learning, not for learning.

It is suggested here that the synergy of formative and summative assessment comes from making use of the same evidence for the two purposes. This can be, as in the Queensland example, where work collected in the portfolio is used to provide feedback to the students at the time it is completed as well as being used later in assessing overall attainment. Here the procedures for using the assessment to help learning are less well defined than in the approach that starts from the formative use. Possibly different emphases are appropriate a different stages of education; the detailed indicators being particularly suited at the primary level where teachers have opportunity to gather evidence frequently but at the same time need more structured help in deciding next steps across the range of subjects they teach.

Synergy also comes from having the same person responsible for using the evidence for both purposes. All assessment involves judgement and will therefore be subject to some error and bias, as the research shows. While this aspect has been given attention in the context of teachers' assessment for summative uses, it no doubt exists in teachers' assessment for formative purposes. Although it is not necessary to be over-concerned about the reliability of assessment for this purpose (because it occurs regularly and the teacher will be able to use feedback to correct for a mistaken judgment) the more carefully the assessment is made, the more value it will have in helping learning. Thus the procedures for ensuring more dependable summative assessment will benefit the formative use and, as noted, the teacher's understanding of the learning goals and the nature of progression in achieving them. For example, experience shows that moderation of teachers' judgements, necessary for external uses of summative assessment can be conducted so that it not only serves a quality control function, but also has an impact on the process of assessment by teachers, having a quality assurance functions as well (ASF, 2004b). This will improve the collection and use of evidence for a formative purposes as well as a summative purpose.

The procedures that will most help both the effectiveness of formative assessment and the reliability of summative assessment are those that involve teachers in planning assessment and developing criteria. Through this involvement they develop ownership of the procedures and criteria and understand the process of assessment, including such matters as what makes an adequate sample of behaviour, as well as the goals and processes of learning. This leads to the position that synergy between formative and summative assessment requires that systems should be designed with these two purposes in mind and should include arrangements for using evidence for both purposes.

References

ASF (2004a) *Working Paper 2 Draft 3 Summative assessment by teachers: evidence from research and its implications for policy and practice.* Assessment Systems for the Future Project

ASF (2004b) Working Paper 1 Draft 9 Assessment Systems for the Future Project

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003) *Assessment for Learning: putting it into practice*. Maidenhead: Open University Press

Carter, C. R. (1997) Assessment: Shifting the responsibility. The Journal of Secondary Gifted Education, 9 (2) Winter 1997/1998 68-75

Cumming, J. and Maxwell, G.S. (2004) Assessment in Australian schools: current practice and trends. *Assessment in Education,* 11 (1) 89 -108

Harlen, W. and Deakin Crick, R. (2003) A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills (EPPI-Centre Review). In *Research Evidence in Education Library.* Issue 2. London: EPPI-Centre, Social Science Research Unit, Institute of Education. Available on the website at: http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/assessment/review_two.htm

Harlen, W. and James, M.J. (1997) Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education* 4 (3) 365 - 380

Harlen, W. and Qualter, A. (2004) *The Teaching of Science in Primary Schools* 4th Edition. London: David Fulton Publishers

Jackson B. (1989) A Comparison Between Computer-Based and Traditional Assessment Tests, and Their Effects on Pupil Learning and Scoring, *School Science Review,* **69,** 809-815.

Kumar D. (1993) Effect of HyperCard and Traditional Performance Assessment Methods on Expert-Novice Chemistry Problem Solving, Annual Meeting of the National Association for Research in Science Teaching., Atlanta GA.

Masters, G. and Forster, M. (1996) *Progress Maps.* Victoria, Australia: Australian Council for Educational Research

Maxwell, G. S. (2004) Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland. Paper presented at the third Conference of the Association of Commonwealth Examination and Assessment Boards, March 2004, Nadi, Fiji.

National Council for Educational Technology (NCET) (1994) *Integrated learning Systems: A report of the pilot evaluation of ILS in the UK.* Coventry: NCET

Osmundson E. Chung G. Herl H. & Klein D. (1999) *Knowledge Mapping in the Classroom: A Tool for Examining the Development of Students' Conceptual Understandings,*

Schacter J. Herl, H.E., Chung, G.K.W.K., O'Neil, H.F.O., Dennis, R.,& Lee, J.J (1997) Feasibility of a Web-Based Assessment of Problem Solving, Annual Meeting of the American Educational Research Association, Chicago.

Wilson, M. 1990. Measurement of developmental levels. In *International Encyclopedia of Education: Research and studies. Supplementary volume,* eds. T. Husen and T. N. Postlethwaite. Oxford: Pergamon Press.

Wilson, M., C. Kennedy, & K. Draney. 2004. GradeMap (Version 4.0) [computer program]. Berkeley: University of California, BEAR Center.